8 Methods in Short-Term Climate Prediction

1.climo, 2.persist, 3.OCN, 4. Local regress, 5 Non-local regress, 6 Composites, 7 regress at pattern level, 8 NWP, 9 Consolidation, 10 Other method, 11 Methods not used, App I, App II

The purpose of this chapter is to list the more common accepted methods used in short-term climate prediction, explain how they are designed, how they are supposed to work, what level of skill can be expected and the references to find more about them. The emphasis is on methodology but aspects of verification and cross validation will be mentioned as well. Most methods will be accompanied by an example. We will also mention some of the less common methods, but with less detail. We even list some methods that are not used, to delineate which are acceptable and which are not. Section 8.1 - 8.6 and 8.8 are easy to read, but 8.7 and 8.9 are more difficult.

It will become clear by the end of the climatology section (8.1), that only the departure from climatology, the so-called anomalies[1], are considered worthy forecast targets. The climatology itself, including such empirically established facts as 'days are warmer than nights', and 'winters are colder than summer', is considered too obvious to be a forecast target. This is not to say that a quantitative explanation of the earth's climate, including daily and annual cycle, is easy. But in professionally honest verification no points are given for forecasting a correct climatology. This chapter is thus about forecasting aspects of the geophysical system that are not so obvious and more difficult. The daily and annual cycle are periodic variations controlled by external forcings such as the solar heating. Implicit in identifying a periodic phenomenon as such is that the forecast of the phenomenon is easy out to infinity. This explains a widespread search for 'cycles' in early meteorological research, but very little has been found other than the obvious

[1] Anomalies, defined as a departure from climatology, have a long history in meteorology. The use of anomalies is very common in modern climate diagnostic studies but it has not always been that way. In the mid 19th century Buys Ballot successfully temporarily championed plotting surface pressure anomalies on weather map so as to make observations taken at different elevations more comparable. The surface pressure anomaly was much later replaced by mean sea-level pressure. The anomaly concept continued mainly in connection with expressing long range weather forecast. In short range weather forecasts the use of anomalies is rare, even today.

daily and annual cycles. By removing a climatology that accounts for daily and annual variations we in effect remove the known easy periodic part of the system.

8.1 Climatology

In the absence of any other information climatology is the best information available. As many travelers can attest, somebody visiting an unfamiliar location 6 months from now is well served by inspecting climatological tables. Climatology is usually defined as a 30 year mean, currently that would be over the 1971-2000 era, while before 2001 it was the 1961-1990 mean. There is nothing magic about 30 years, but it came about (almost 100 years ago) as a long debated compromise in the standards adopted by the WMO, balancing a desire to have much longer averages on the one hand (which would yield better estimates of the mean, median or 'expected value' in a constant climate), and the practicality that most of the homogeneous station records are short on the other. Note that official WMO climatology is like a discontinuous moving window.

Climatology, as a 30 year mean or otherwise, is often used as the 'control forecast' in verification, meaning that any forecast method claiming to be useful should beat climatology in terms of an accuracy attribute A, like rms error (Murphy and Epstein 1989). Skill is usually defined as

$$SS = \frac{A_c - A_m}{A_c - A_p}$$

where the subscripts c, p and m refer to climatology (or control more generally), the perfect forecast and the method to be verified respectively. For rms error the attribute A_p is zero and one wants A_m to be smaller than A_c, and the upper limit of skill is 1 (when $A_m = 0$). This is also the exact philosophy underlying the anomaly correlation, see appendix in chapter 2[2]. Climatology itself then is the base-line forecast with zero skill by construction. In essence one needs to design a verification system that does not give any credit for forecasting climatology thus leaving only departures from climatology a legitimate forecast target. A downside of this strict professional attitude is that if there is no skill by that standard there is a tendency at Weather Services to not issue forecasts at all. After all why publish a forecast without skill? However, many users, especially casual users, need to be reminded that 30 year means are the fall-back 'forecast' in that case. Not every user has ready access to this information or understands the situation they are in when Weather Services claim they cannot make forecasts with skill[3]. The 30 year mean is still better than a random guess.

One can obviously use a 30 year data set to learn much more than just the mean. Standard deviation, extremes (records), probability of exceedence for certain thresholds etc can be estimated as well. One can define climatology as a pdf, rather than an expected value only - this makes sense also because the profession is moving in the direction of probability forecasts and in that context the control forecast needs to be a pdf as well. Keep in mind that higher order quantities may need many more than 30 years for an accurate estimate, but in a changing climate the expected value loses relevance when evaluated from data extending too far back. This situation is too complicated for anything as simple as WMO standards.

Climatology forecasts are completely local in space. Information from other locations is not used. Creating climatology from neighboring measurements is tricky, especially near orography or other geographical boundaries (land-sea).

Quantitative application of climatology forecasts does require accurate measurements to

---

[2]A positive anomaly correlation indicates that a properly damped forecast anomaly has lower rmse when verified against the observed anomaly than the zero anomaly climatology forecast. Proper damping in this context means that forecast anomalies are multiplied by the anomaly correlation times the ratio of observed to forecast standard deviations.

[3]Additional confusion: Other providers of forecasts may claim skill even if the Weather Service believes there is no skill.

be taken at the location of interest for many years. In that sense even the most trivial of trivial forecasts does require a sustained effort and does not come for free. And even with all the data at hand, creating climatology from a long and accurately measured time series also has its difficulties. Some authors have recommended smoothing of the data by allowing only a few harmonics, say the annual, semi-annual terms plus maybe one or two more harmonics (Epstein 1981, Trenberth 1984 Schemm et al(1997)). In doing so the estimate of the standard deviation around the filtered climatology also gets adjusted. These are choices to be made by researchers and practioners.

By necessity, the climatological control forecasts that go along with forecasts issued in real time refer to past climatology (like 1971-2000). For forecasts made retroactively (a model can be run after the fact say over 1981-2003 for the sake of argument, see Saha et al(2006)), the climatology over 1981-2003 is available. The 1981-2003 climatology is centered in time relative to the retrospective forecast data set, thus creating a climatological control forecast that differs from 1971-2000. This creates some difficulties in comparison of skill estimates and interpretation because the climate may change and the control for verification becomes an issue.

The non-constancy of climatology will be treated in the section on OCN.

8.2 Persistence

One notch up in complexity and usually skill is a simple method called persistence. In words this means that current conditions are the forecast for a later time. For instance, today's maximum temperature is 27C, so the forecast for tomorrow is also 27C. This is essentially a lazy person's forecast with no great expense involved, but because the atmosphere (and especially the ocean) vary slowly, except in special and somewhat rare circumstances, this sort of forecast usually has higher skill than climatology. There is a large collection of related persistence forecasts:

1) persistence of a previous value, i.e. $f(t)$ is the forecast for $f(t+\Delta t)$

2) persistence of the previous anomaly, i.e. $f'(t) = f(t) - clim(t)$ is a forecast for $f'(t+\Delta t) = f(t+\Delta t) - clim(t+\Delta t)$.

3) damped persistence, i.e. $\mathbf{a}\, f'(t)$, $0<=\mathbf{a}<=1$, is a forecast for $f'(t+\Delta t)$

If the climatology does not change over the Δt interval, methods 1) and 2) boil down to the same thing, although the explicit reference to climatology in method 2 may be a helpful reminder to some users. Persistence of the anomaly, method 2), makes clear that the skill of this forecast, when measured by correlation, in the long run equals the autocorrelation:

$$\rho = \Sigma f'(t) \quad f'(t+\Delta t) / \sqrt{\ } \{ \Sigma f'(t) \quad f'(t) \quad \Sigma f'(t+\Delta t) \quad f'(t+\Delta t)\}, \qquad (8.1)$$

where summation is over many t (but stratified by time of year). Hence, if the process we are forecasting has a long time scale, skill will be very high for small Δt due to persistence, and may be hard to improve upon for any of the more daring methods. The 3rd version, damped persistence, uses the a-priori known skill in formulating the forecast. In order to minimize the rmse of the forecast the factor **a** in $\mathbf{a}f'(t)$ should be the autocorrelation times the ratio of the observed to forecast standard deviations.

Clearly the 3rd method is better that the first. This is especially true for large Δt, say months, when the mean and the standard deviation need to be adjusted, and can be adjusted if measurements have been taken for many years.

Although a persistence forecast does not sound exciting, there have been numerous studies about persistence of both surface weather elements and upper air data (Namias 1952, Dickson 1967, Van den Dool et al 1986). This is because the degree of persistence varies with location, time of year, and the element studied. A full understanding of such variations is a big challenge. Some oceanographic and atmospheric processes can be closely approximated by a first order Markov process, so the study of persistence is a study of the whole power spectrum in one number, the autocorrelation at one particular lag (Gilman et al 1963).

Fig. 8.1. shows an example of the annual cycle of persistence of surface air temperature over the US. The calculations are done with standardized data for many years, placing all 102

Climate Divisions on equal footing, then aggregated by summing in space, see Van den Dool et al (1986) for details. We find that the monthly mean temperature correlates at about +0.2 from one month to the next (say June to July, called lag 1) and between 0 and +0.15 at lag 2, which would, for example, be forecasting July with May temperature data. Although the correlations are modest, they are positive without exception and the results are undoubtedly statistically significantly non-zero. So in general a month that is above (below) the mean tends to be followed more often than not by positive (negative) anomalies. Thus, there is a tendency for the anomaly to repeat itself. One should note that persistence has a maximum both in late winter and late summer, with lower values in between. The two maxima and minima are typical for the skill of many short term climate prediction schemes, and in great contrast to NWP, which has a single skill maximum in winter (around the time the degrees of freedom in the atmosphere are lowest) and a single minimum in summer when spatial scales of weather systems are smallest. An attempt to explain the annual cycle in persistence was given in Van den Dool(1983).

Local persistence is much higher along some shorelines. For instance San Diego (USA) and Den Helder (The Netherlands) may have 0.6 - 0.7 month-to-month correlation at certain times of the year due to the inertia of nearby SST anomalies (Van den Dool and Nap 1985). Schemes that use nearby SST instead of a previous air temperature suggest themselves. One may suspect local effects of this sort near lakes, snow fields, soil moisture, anything at the lower boundary that is capable of extending persistence in anomalies. One should also note that if 2 weeks were the limit of predictability (l.o.p.), as suggested by many fraternal twin NWP experiments, there should not have been any correlation in Fig.8.1 at all for lag 2. The non-zero correlation at lag 2 is thus evidence that the l.o.p. can be punctured by the simplest of means. In some cases it is obvious that this is the result of long lived anomalies in the lower boundary conditions.

One can introduce persistence as a control forecast, and it would be a (much) tougher control than climatology. A strong argument in favor of measuring skill as an improvement over persistence is that the time rate of change is the essence of forecasting, especially if one thinks of the forecast problem in terms of the basic equations with a time derivative on the left hand side. A zero time derivative is the same as persistence (of type 1)), so if we cannot beat persistence the most essential aspect of knowledge about the future is missing.

Just as with climatology, the persistence forecast is purely local, as information of neighboring areas is not used. To apply persistence and especially damped persistence reliable measurements are required, and preferably with more than 30 years of data since calculating the damping factor requires estimates of higher order statistics.

When calculating the damping factor **a** as a function of Δt from data, one may, in some rare cases, discover negative values. This would obviously happen if an oscillatory mode dominates the data and/or the process is clearly not a first order Markov process. This situation will be covered more extensively when describing regression methods below. But negative **a** indicates a number of possibilities for obvious extensions, most notably the use of several previous values e.g. use of both f(t), and f(t-Δt) for a forecast valid at t+Δt .

A source of some debate and confusion is exactly what quantity is, or should be, persisted. Along with the aforementioned methods 1-3, should the variable be the very latest value, or the latest weekly mean, the monthly or seasonal mean?? Should the length of the time mean of predictor and predictand be the same? And when a (damped) anomaly is persisted, with respect to which climatology is the anomaly taken? Some of the questions raised in the Climatology section (8.1) present themselves again. For instance, in Fig.8.1, we calculated persistence in anomalies that are departures from the 1932-2000 mean, but in a real time forecast during any of these years that would have been impossible since the long term mean is not known. There are no strict answers, just choices for the researcher and practitioner, choices that need to be justified in some specific context. How about persisting the average over the last 10 years? This will be discussed the next section.

8.3 Optimal Climate Normals

The definition of climatology as a 30 year mean, adopted by WMO in the early 20th century, was never fully accepted. This is primarily because the earth's climate is evidently not constant, especially in temperature. In a slowly varying climate (as opposed to a stationary climate) the average over the last K years may be a better estimate of the upcoming expected value than a longer term mean. OCN stands for optimum climate normals, where the optimum refers to the value of K that minimizes U(K)

$$U(K)=\sum_{j}\{T_K(j)-T(j+1)\}^2,$$

where T(j) is the temperature in year j, and $T_K(j)$ is defined as

$$T_K(j) \equiv \sum_{j'=j-K+1}^{j} T(j')\,/K$$

There is no analytical approach to minimize U. Instead all values of K are tried. Evaluation of U(K) for each K for temperature over the US (Court 1967/68 ; Huang et al 1996, Wilks 1996) has revealed that U is minimum for a value of K on the order of 10 years. There is regional and seasonal variation in K although such variation may not be known accurately enough. The U(K) function is often very flat, such that the optimum K is not that much better than neighboring K values, or has several minima. Moreover, the function U can also be evaluated in terms of absolute error, or in terms of correlation (Lamb and Changnon 1981), leading to similar, but not identical, K. The main point is that K is considerably shorter than 30 years for temperature: U(K) < U(30). Moreover the official 30 year normal is aging until the next update (once every 10 years at best[4]), so that U(K) < U(30) < U(30-fixed period). There are thus two reasons as to why OCN is a better forecast (lower rmse) than the official 30 year normal: a) K<30, and b) instant update of the K year average.

[4]WMO has had only 1931-1960, 1961-1990 normals. But many countries update every 10 years, so 1971-2000 is in effect as normal in 2006.

For certain well defined processes, one can derive the value of K. For instance in a stationary climate K=∞ (or as close as one can practically be to ∞). For a red noise process K=1. For a linear trend (and negligible noise superimposed) K=1 also. The empirical value of K is another synopsis of the power spectrum into one number.

The raw forecast anomaly due to the OCN forecast methods is simply the difference of $T_K(j)$ and the official 30 year climatology. This difference, for K=10, is evaluated continuously for the US, and is shown for each rolling season at http://www.cpc.ncep.noaa.gov/products/people/wd51hd/ocn.html . Fig.8.2 shows the OCN temperature maps for the 4 canonical seasons, as of Jan 2006. The temperature averaged over the last 10 years is, more often than not, higher than the 1971-2000 normals. This is especially so in the interior SW USA where all seasons appear to warm and the difference is approaching one standard deviation (of seasonal mean data around the 71-2000 mean), a huge shift. The only exception to the general warming is an area in the northern plains during spring into summer when temperatures have become lower. Similar maps for Precipitation (using K=15) are also maintained at the aforementioned website, but the shifts are much less impressive and not nearly as useful for prediction.

OCN has been used internally by certain industries for a very long time to provide a rational basis for calculating prices of such commodities as electricity, heating oil etc. While OCN is a baseline forecast for climate in the future, OCN had also been used informally in seasonal forecasts at the Climate Analysis Center (now Climate Prediction Center), and at the UK Met Office (Gilchrist 1986). Changes in the earth's climate, in temperature especially, have become quite evident in the last 20 years. This has made OCN, or other tools that assess 'trends', far more important for the seasonal forecast than imagined previously. Since late 1994, OCN has been used in a formal way in the US seasonal forecasts, see chapter 9. In 1994 we had no idea how important this would be.

OCN is a local method, requiring data only from the site where a prediction is needed.

One can look upon OCN either as an amended version of climatology (one that needs continual study and adjustment) or persistence of the average anomaly of the last 10 years. K=10 may not stay optimal of course. One may feel OCN is really a forecast for the next 10 years averaged. When verified in that fashion the correlation goes up from 0.3 for single seasons to at least 0.6 for 10 year averaged seasonal anomaly.

Interdecadal climate variation has been reported in many variables, such as net seasonal Atlantic hurricane activity (Chelliah and Bell 2004) and drought (McGabe et al 2004). To what extent these OCN like variations can be described as one phenomenon remains a subject of study. CPC predicts the net seasonal hurricane activity by mainly empirical means and in addition to interdecadal variations ('hurricane OCN') ENSO plays a role.

8.4 Local Regression

The persistence methods, especially the 3rd version, are special cases in the category of local regression. I.e. if f(t) is the predictor and g(t) is the predictand, we seek to determine the coefficients a and c in the expression

$$g^F(t+\Delta t) = a * f(t) + c \qquad (8.2)$$

where g(t) and f(t) are data sets co-located in space. The coefficients a and c can be derived through a regression based on sufficient observations of $g(t+\Delta t)$ and f(t). f and g are time lagged, but the number of time levels is the same for f and g. In the persistence methods discussed above, f and g were the same data set and for g=f Eq(8.2) is autoregression. The intercept c would be zero if f and g are anomalies, and the regression coefficients are calculated over the cases that go into forming the climatology.

Fig. 8.3 shows an example. The soil moisture anomaly at the end of the month is the predictor and the predictand is the co-located monthly mean temperature in the next month (lag 1) and 2 months later (lag 2). Soil moisture is not generally observed, but is calculated by a simple physical soil model forced by observed weather elements, integrated forward from 1931 to the present (Huang et al 1996). The rationale for the exercise is that dry (wet) soil leads to decreased (increased) evaporation[5] and thus increased (decreased) temperature. This effect should only be seen when there is enough incoming solar radiation, since it is the latter that is used either for evaporation or heating the air by the sensible heat flux. Fig. 8.3, as was Fig.8.1, is an aggregate result for all of the 102 US 'super' Climate Divisions combined. In general, one can indeed see a negative correlation between soil moisture and subsequent temperature, mainly from April through October, as would be expected. The lag 2 results are similar, but a little weaker. July is more often than not colder than average after May has ended with wetter than average soil. The effect of soil moisture may explain in part why temperature itself is persistent in summer (Huang and Van den Dool 1993).

While the correlation is negative in Fig. 8.3, the skill of a forecasting scheme based on soil moisture is the absolute value of the shown correlation as long as results hold up perfectly on independent data. Because we used a lot of data (70 years, 102 locations) that may be very nearly the case here. The degradation of a full sample correlation (ρ_{fs}) in a CV-1-out test, is given by (Barnston and Van den Dool(1993))

$$\rho_{cv} = \frac{L-1}{L} \rho_{fs} - \frac{1}{\rho_{fs} (L-1)} \quad (8.3)$$

where ρ_{cv} is the correlation after cross validation is done, and sample size L = N * M, the product of N, the number of degrees of freedom in space (which is 5 to 15 for temperature in the US as evaluated from Climate Divisions (Huang et al(1996)) and M (the number of years if year to year autocorrelation is small). Shown in Figs 8.1 and 8.3 are full sample correlation (ρ_{fs}) and ρ_{cv} as per Eq 8.3 would be the correlation to expect on independent data. The two terms in (8.3) combine shrinkage of the correlation as well as degeneracy of the CV 1 out procedure. For local correlation, L is at most 70, and a full sample correlation of 0.30 for example should not be

[5]By evaporation we mean the combined evaporation from all sources, i.e. from bare soil, bodies of water, water on the canopy as well as transpiration by green vegetation during daylight hours.

expected to yield more than 0.24 when applied as a regression forecast. The spatially aggregated correlations are expected to hold up better on independent data because L is much larger.

Aggregation in space of local results in Fig.8.1 and 8.3 makes sense only because we expect somewhat similar effects everywhere, and the aggregation greatly reduces the noise. Aggregation would make no sense at all if soil moisture correlated positively (negatively) with subsequent temperature in one half (the other half) of the domain. Fig.8.4 suggests some structure in the time lagged soil moisture - temperature correlation, but the sign of the lagged correlation is negative nearly everywhere. We have chosen spring in Fig.9.4 because this is a very interesting season. The state of soil moisture at the end of February has only a small impact in March (upper right), but suddenly grows in importance in April (lower left) in the SE USA. This may be related to the greening of plants as the season advances - after all, most of the evaporation comes through the vegetation. Note also that the simultaneous correlation (upper left) is +ve in a few areas. This happens in winter in areas where precipitation is associated with warm air advection. But correlations are negative nearly everywhere at positive lag.

No one stops us from considering a local regression involving yet another data set, or several others:

$$g^F(t+\Delta t) = a\ f(t) + b\ \ h(t) + c \qquad (8.2a)$$

where g, f and h could be co-located temperature, soil moisture and snow depth for instance. The problem with having to estimate three or more coefficients (as a function of space, month and Δt) is having enough observations to make reliable estimates. In short term climate prediction we rarely have a lot of data so prudence is very important. It helps to have a physical basis for picking a certain predictor. Trying out every predictor under the sun is a certain disaster.

A special case of (8.2a) would be:

$$f^F(t+\Delta t) = a\ f(t) + b\ \ f(t - n\Delta t) + c \qquad (8.2b)$$

where f is forecast from several previous values of itself. It does happen in some rare instances that a, when determined from regression, goes negative for certain values of Δt. As a curiosity

Namias(1952) and Dickson(1967) would report 'anti-persistence', i.e a tendency for the anomaly to change sign at the chosen lag. If reliable, this would indicate a sufficient departure from a first order autoregressive to admit one or more previous values. For instance if f(t) were a sine wave the best prediction would be to use the latest value and another value one quarter of the period earlier. A great example, perhaps the only one, is the Quasi Biennial Oscillation.

Fig.8.5 shows the time series of the zonal mean zonal wind at 30mb for the period 1948-present. The somewhat cyclical variation, known as the QBO in the stratosphere, invites a scheme like (8.2b) for its prediction, where $n^*\Delta t$ should be about one quarter of the period[6]. Prediction of the QBO has been deemed important for seasonal hurricane activity forecasts (Gray et al 1994). The forecasts of the QBO can be made using the naked eye, i.e. it is nearly obvious for a few seasons out. However, the period ('Quasi-Biennial') is not exactly constant and the amplitude varies as well, so forecast schemes of the QBO are not as successful (or trivial) as a forecast of the ever repeating annual cycle or the tides in the ocean. The QBO, in spite of its near regularity, is not known to be linked to external forcing, but is caused by dynamics internal to the atmosphere (Holton and Lindzen 1972; Plumb 1977). The phenomenon was discovered in the late 1950's (Reed et al 1961). The QBO is by far the most predictable unforced signal in the atmosphere, very much beyond the 2 week limit of predictability, but its impact on surface weather and climate remains largely unclear. Moreover, there may be weaker biennial signals of a different origin near the surface (Barry and Carleton 2001).

Note that at the beginning of the graph in Fig. 8.5 the QBO appears absent. In the early years, 1948 to mid-fifties, there were no observations that high up in the stratosphere to be assimilated in the Reanalysis - in reality there probably was a QBO even though Fig.8.5 does not show one. The Reanalysis has yielded a beautiful and accessible data set, but one can never take data for granted.

[6]An alternative is to use the zonal wind simultaneously at various levels in the vertical because the QBO signal is known to propagate downward.

The QBO may have high -ve correlation at a lag of about 13 months, but closer to the ground anti-persistence in wind, temperature or pressure is always small and may be sampling error. If anti-persistence were common, persistence would not be a recommended control forecast, and would not be an improvement in skill relative to climatology. There are only a few locations near the storm track where the passage of cyclones is regular enough to suggest a modest negative autocorrelation in wind and pressure after a few days. Van den Dool and Livezey(1984) report on the spatial distribution of persistence of monthly 700mb height anomalies. The lag 1 autocorrelation is generally positive, especially at low latitudes, but there is a puzzling case of month-to-month anti-persistence in spring in the mid-latitude Atlantic troposphere. Synoptic experience also suggests major changes in the Atlantic area can take place in that season. Black et al (2005) have tried to explain this by up-and-down propagation of NAO like signals into the stratosphere and back, and linking spring anti-persistence near the surface to the process of 'final warming' of the stratosphere.

One could give the procedure of local regression an analytical twist by assuming a 'model'. For instance in Eq (8.2), when written as $f^F(t+\Delta t) = \mathbf{a} * f(t) + c$, the value of $\mathbf{a}(\Delta t)$ could be either calculated for each Δt, or we calculate $\mathbf{a}(\Delta t)$ for only one specific value of Δt, but generalize the resulting **a** by assuming how $\mathbf{a}(n*\Delta t)$ relates to $\mathbf{a}(\Delta t)$. For instance, in a first order Markov chain the autocorrelation drops off theoretically as $\rho(n*\Delta t) = \rho^n(\Delta t)$[7]. One Δt calculation suffices, or one could smooth the results a bit from an evaluation of coefficient **a** for several values of the time lag. This semi-analytical approach will be important later on in the discussion of LIM.

8.5 Non-Local Regression and ENSO

It is methodologically obvious how to pursue non-local regression. In simplest form we

[7]Lorenz(1973) noted that autocorrelation showing positive departures from a 1st order Markov process at longer lags is basic empirical evidence of longer range predictability.

have

$$g^F(s_1, t+\Delta t) = a * f(s_2, t) + c \qquad (8.4)$$

where we seek to forecast g at location s_1 from the predictor f at location s_2. Position s_1 may not be arbitrary, since it is the location for which a forecast is desired, but the choice of s_2 may be more difficult. There has to be a good justification and/or thorough statistical testing. Perhaps the best known example of a relationship that fits Eq (8.4) is given in Fig.8.6 which shows the correlation between Nino34 (or Darwin pressure) in SON and the seasonal mean temperature and precipitation over the US in the subsequent JFM period. The calculation is based on data during 1931-2004. The ENSO related teleconnection is thought to explain this pattern. While teleconnections are near-simultaneous, anomalies in tropical indices like Nino34 are long lived, and the lagged correlation displayed in Fig.8.6 is thus about forecasting a teleconnection several months ahead of time. In Fig 4.2 we showed a one point teleconnection pattern for 500mb height in JFM which may accompany the surface weather elements shown in Fig.8.6. (Figs. 7.4 and 7.5 give a specific example for February 1998, which conforms well to the canonical pattern during ENSO.) There is no such 500 mb pattern teleconnection during SON itself. But communication with the tropics opens up when the westerlies advance farther to the south with season (Opsteegh and Van den Dool 1980) and the teleconnection emerges.

What is especially noteworthy about Fig.8.6 is that the ENSO impact on precipitation is about as large (or larger) as the impact on temperature. It is highly unusual for mid-latitude forecasts of precipitation to have skill anywhere near the skill level of temperature. (In fact we did not show the companion results in Fig.8.1 and 8.3 for precipitation because the correlations are indistinguishable from zero.) Here is the one exception. During ENSO warm events the southern US, and especially the Southeast and interior Southwest are likely to be wet. In South Florida the correlation is better than 0.60. Note also the gradient in the correlation between Kentucky and Florida, indicative of a change in storm track across the Southeast US. The temperature signal (upper panel in Fig 8.6) is for positive anomalies across the north and negative anomalies in the

Southeast when Nino34 is positive. Maps based on Nino34 and SLP at Darwin as predictor during SON look very similar. Darwin pressure is one of two pieces of input to the SOI, and the SOI is closely related to the ocean's SST. That the measurement of pressure at a single location in Australia (or SST in a small area) would have such far reaching prognostic consequences is among the more exciting aspects of our profession. It also seems odd that we can make such statements based on very simple means, given that forecasting the weather a few days out requires detailed input initial data on a global domain. How can short-term climate prediction be so simple?

The impact of ENSO on climate far away was not known very well before about 1980, in part because global data sets had not been available in real time before 1980. Douglas and Engelhart(1981) reported a high correlation between autumn precip on islands in the Pacific and winter rainfall over Florida, much the same as shown in Fig.8.6. Madden and Van Loon(1981) reported on simultaneous global effects of a measure of the Southern Oscillation and pressure and temperature elsewhere, where-ever there was data. The strong 1982/83 event, along with the availability of global data (satellites as well), prompted interest that has remained high ever since. The ENSO teleconnection research is often about simultaneous relationships, the assumption being that because of high persistence in the tropics any simultaneous relationship carries over at lag.

In spite of the long distance ENSO impact over land, even more so in Australia, Indonesia and South America than over North America, one should not lose sight that the correlations (over the US) are still modest with absolute value exceeding 0.4 in only limited areas and mainly in winter. Most of the area shown in Fig.8.6 has correlations less than +/-0.2, which means that most of the US does not have a significant linear correlation with ENSO. Mason and Goddard(2001) come to the same sobering conclusion for the entire globe. Linear correlation has obvious drawbacks - more on ENSO will be discussed in the section on composites.

Except for links to ENSO, there may not be that many non-local simple prognostic relations that fit Eq (8.4). However, if we relax the definition and allow g(s1,t) and f(s2,t) to be the sine and cosine components of moving waves, EWP (which could be named Lagrangian persistence) may be looked upon as the complex version of Eq (8.4), i.e.

$$g^F(s_1, t+\Delta t) = a * f(s_2, t) + b * g(s_1, t)$$

$$f^F(s_1, t+\Delta t) = c * f(s_2, t) + d * g(s_1, t)$$

which, as shown in Chapter 3, when executed wave by wave, leads to the expressions for empirical wave propagation. For unfiltered daily data EWP has some skill, but no prognostically useful phase propagation is known to exist for filtered monthly or seasonal mean data.

For those seeking more references on methods in non-local regression, there is a body of literature on Model Output Statistics beginning with Glahn and Lowry(1972).

8.6 Composites

A composite is an operation on a subset of the available data, which is conditioned to satisfy some criterion. For example, one can calculate the mean winter temperature in Washington DC for those years when DJF Nino34 was at least 0.5C above the mean. In the same way, one can make probability statements: during years when Nino34 SST was at least 0.5C below the mean, the precipitation at some locale in the Southern US was 8 out of 10 times in the below normal tercile. Such composites should be broadly consistent with Fig.8.6 but allow for a) an evaluation of asymmetry (non-linearity) in the relationship between Nino34 and remote weather and climate, and b) an understandable user friendly forecast. Is the impact of a La Nina the exact opposite of an El Nino as assumed in linear correlation? If a cold event is expected for next winter, the statement that "8 out of 10 previous warm events were particularly dry" is clear to even the person in the street. Of course one can make composites based on any criterion. The reason ENSO is used here as an example is because Nino34 is i) quite predictable months ahead of time at most times of the year, and ii) has discernable influence. Composites as a technique are widely used for diagnostic research. One may study the mean weather conditions given that the

PNA or NAO are simultaneously far above normal. In many places in Europe and North America, the PNA and NAO have a larger influence than ENSO but because prediction of NAO and PNA is far less successful, an NAO composite is mainly a diagnostic tool. In Chapter 4, we used compositing to study asymmetry in the relationship between base points near Greenland and Europe. Research into the skew, asymmetry and or non-linearity of ENSO includes work by Burgers and Stephenson(1999), Montroy et al(1998), Hoerling et al(1997), Lin and Derome(2004), and Wu et al (2005).

ENSO composites have been a standard tool in long range forecasting since Ropelewski and Halpert(1986, 1987, 1989) and Halpert and Ropelewski(1992) made their famous display of world wide short term climate anomalies related to the Southern Oscillation. However, the details of the compositing method have changed since then. There does not appear to be a settled method. At the CDC website http://www.cdc.noaa.gov/index.html one can make composites for the US 'on demand' in an interactive database&graphics system. On the CPC website one can find a fixed set of US composites used by its forecasters at http://www.cpc.ncep.noaa.gov/products/precip/CWlink/ENSO/total.html . However, even here the number of options is large, due to trend adjustment considerations (adding OCN into the composite; see Higgins et al 2004). The composites used for the famous 97/98 ENSO winter can be found in Barnston et al(1999). More on composites can be found in Wolter et al(1999) and Cayan et al(1999).

Dangers with composites are obvious. Since they are based on fewer cases (than a linear regression using all data), the result is noisier and may be over-interpreted. Sampling variability may be confused with true non-linearity. Also, minute changes in the threshold may change the number of cases (and the results) significantly. The criterion for ENSO is debatable, but in recent years NOAA has adopted a 0.5C Nino34 anomaly criterion in three month running mean Nino34

SST for its real time guidance[8]. The advantage and simplicity of composites is quickly lost when more than one factor or other considerations enter the final forecast.

8.7 Regression on the Pattern Level

Most empirical methods in short-term climate prediction are nowadays based on multiple linear regression 'on the pattern level'. A primitive example is as follows. Suppose we have two data sets, f(s,t) called the Predictor, and g(s,t) called the Predictand. One can perform two stand alone EOF analyses of f and g, and then do the regression between the time series of the leading modes in the predictand and predictor data sets. Klein and Walsh(1984) made an in depth comparison of regression between EOF mode time series on the one hand and regression between the original data at gridpoints on the other - this was in the context of 'specification' (as discussed for instance in Ch 7.3). Using modes is efficient, and cuts down on endless choices, but it may not always help the skill.

For a more general approach we first discuss the time lagged covariance matrix.

8.7.1 The time lagged covariance matrix

When we have two data sets, f(s,t) called Predictor, and g(s,t) called Predictand, one can define the elements of the time lagged covariance matrix C_{fg} as

$$c_{ij} = \sum_{1}^{n_t} f(s_i,t)\ g(s_j,t+\tau)/n_t \qquad (8.5)$$

where n_t is the number of time level, a time mean of f and g was removed, and τ is the time lag. C, non-square in general, thus contains the covariance between the predictor at any place in its domain, and the predictand anywhere in its domain - local as well as non-local. (From the time lag in g our intention is clear: to predict g from f. However, some analyses below (CCA, SVD) do not go beyond establishing associations between f and g, leaving in the middle who predicts

---

[8]For historical research purposes warm or cold events should have at least 5 successive rolling seasons in excess of the 0.5 criterion.

whom. Most texts on SVD and CCA thus do not show a time lag)

Associated with c_{ij} there is also a correlation

$\rho_{ij} = n_t * c_{ij} /\mathrm{sqrt}(\Sigma\, f^2(s_i,t)\ \Sigma\, g^2(s_j,t+\tau)\ )$ (8.5a)

If g and f were the same data set, and the time lag is zero, C would be the square Q, as per Eq (2.14). Along with C_{fg} we also need Q_f and Q_g below. ($Q_f = C_{ff}(\tau=0)$). Given how Q was manipulated to calculate EOF (presented in chapter 5 as 'self prediction') one may surmise that C can be used to relate patterns&time series in the predictor field to patterns&time series in the predictand field. Indeed, C, in its various renditions depending on prefiltering, truncation, orthogonality constraints, organization of input data sets, etc, is among the most studied in short term climate prediction. Instead of the role played by the notion explained variance (EV) in EOFs, the target of calculating coupled patterns/time series is often in 'explaining' the co-variance of f and g. Because covariance can be negative, the target is often taken to be 'squared covariance' (SC). i.e. the fraction of $\Sigma c_{ij}^{\,2}$, where summation is over all i and j, that can be explained by 1, 2 or m coupled 'modes'.

Without any truncation or constraint C is set up to create any imaginable regression between f and g, so as to minimize the rmse of the prediction of g, on the dependent data that is used to compute C. Here lies a very significant problem. With so many predictors f(s,t), it is hard to avoid overfitting[9]. C contains the correlation of everything with everything. The overfit is combated by severe truncation at the pattern level. This reduces the subjective nature of choosing predictors.

Somewhere in C also lie the methods we already discussed before, like local persistence and local and non-local regression. The reason to present these simpler methods separately and upfront is twofold. First we may easily lose local effects when applying truncation at the pattern level, i.e the very high persistence in temperature in San Diego California would not make it into a

---

[9]Standard texts on regression should be consulted to find methods of exploratory regression that can avoid overfit in most cases.

pattern method until hundreds of modes are admitted. Secondly, C is calculated without any physical intuition. The local effects approach can more easily be defended on physical grounds.

8.7.2 CCA, SVD and EOT2

In chapter 5 we presented EOFs of the data set f(s,t) as:

$$f(s,t) = \sum_{m=1}^{M_f} \alpha_m(t)\, e_m(s) \qquad (8.6)$$

where both the time series and spatial patterns are orthogonal. Eq (8.6) still gives a complete representation of f as long as either the time series or the spatial patterns are orthogonal, and M_f is large enough. Likewise we have for the predictand:

$$g(s,t+\tau) = \sum_{m=1}^{M_g} \beta_m(t+\tau) d_m(s) \qquad (8.6a)$$

Coupling the modes among the two data sets f and g, which have the same number of time levels but possibly different spatial domains (also $M_f \neq M_g$), will be discussed below in terms of the properties of $\alpha_m(t)$ and $\beta_m(t+\tau)$ and $d_m(s)$ and $e_m(s)$ respectively. In any of the methods below orthogonality is maintained in either time or space (not both), so the coupled modes allow projection of future data and/or partial rebuilding of f and g themselves with a set of modes, and the notion explained variance (not optimal obviously) within each data set still applies.

The plain distinguishing feature of Canonical Correlation Analysis (CCA) is that the correlation of $\alpha_m(t)$ and $\beta_m(t+\tau)$, denoted cor(m), is maximized - the modes are ordered such that cor(m)>cor(m+1) for all m. Within each data set we have for CCA

$$\sum \alpha_k(t)\ \alpha_m(t) = 0 \qquad \text{for } k \neq m \qquad \text{(CCA-1)}$$

$$\sum_t \beta_k(t+\tau)\ \beta_m(t+\tau) = 0 \qquad \text{for } k \neq m \qquad \text{(CCA-2)}$$

i.e orthogonal time series, and across the data sets:

$$\sum \alpha_k(t)\ \beta_m(t+\tau) = 0 \qquad \text{for } k \neq m \qquad \text{(CCA-3)}$$

$$\sum \alpha_k(t)\ \beta_m(t+\tau) = \text{cor}(m) \qquad \text{for } k = m \qquad \text{(CCA-3a)}$$

where summation is over time. The cor(m) can be found as the square root of the eigenvalues of the matrix $M = Q_f^{-1} C_{fg} Q_g^{-1} C_{fg}^T$ (or from $Q_g^{-1} C_{fg}^T Q_f^{-1} C_{fg}$). Note that CCA's maps are not orthogonal.

On the other hand, in a method often called singular value decomposition (SVD) the explained SC is maximized. For SVD[10] we have within each data set:

$$\Sigma\, e_k(s)\; e_m(s) \;=\; 0 \qquad \text{for } k \neq m \qquad \text{(SVD-1)}$$

$$\sum_s d_k(s)\, d_m(s) \;=\; 0 \qquad \text{for } k \neq m \qquad \text{(SVD-2)}$$

i.e. orthogonal maps, and across the data sets:

$$\Sigma\, \alpha_k(t)\; \beta_m(t+\tau) \;=\; 0 \qquad \text{for } k \neq m \qquad \text{(SVD-3)}$$

$$\Sigma\, \alpha_k(t)\; \beta_m(t+\tau) \;=\; \sigma(m) \qquad \text{for } k = m \qquad \text{(SVD-3a)}$$

where $\sigma(m)$ is the m'th singular value of C_{fg}. The SC explained by mode m is $\sigma^2(m)$.

Notice the (dis)similarities of SVD and CCA. CCA has orthogonal time series, SVD orthogonal maps. Properties (CCA-1) and (CCA-2) vs (SVD-1) and (SVD-2) appear to be a matter of space-time reversal, but this can not be stated for the 3rd property. The roles of cor(m) and $\sigma(m)$ appear similar. The notion 'SC explained' is sometimes also used for CCA, but does not relate trivially to cor(m). Theoretically it is possible that the first CCA mode describes a perfectly coupled f-g process of infinitesimal amplitude (high cor, low SC).

CCA and SVD are methods to find coupled modes, but they are not quite forecast methods. A regression between the $\alpha_m(t)$ and $\beta_m(t+\tau)$ is needed to forecast $\beta_m(t+\tau)$ given $\alpha_m(t)$.

An easy way of explaining both the idea and the actual application of methods like CCA and SVD to a forecast situation may be to use 'EOT2' - we used EOT in Chapters 4 and 5, but extend it here to 2 data sets. Specifically, we seek the position s_1 in space so that the time series $f(s_1,t)$ explains the most of the variance in the predictand data set g(s,t) at lag τ. I.e. we find i for which U(i) defined as

---

[10] We use the name SVD, even though we agree with Zwiers and Von Storch (1999) that it is unfortunate that the name of the method is confused with a basic matrix operation; they suggest Maximum Covariance Analysis.

$$U(i) = \sum_j \left(\rho^2_{ij} * \sum_t g^2(s_j,t+\tau) / n_t \right) \qquad (8.7)$$

is maximum. Having found s_1 we take $f(s_1,t)$ to be the first mode's time series of *both* f and g expansions, i.e. $f^{explained}(s, t) = a(s_1, s) f(s_1, t)$, and $g^{explained}(s, t+\tau) = b(s_1, s) f(s_1, t)$, where $a(s_1, s)$ is the regression coefficient to predict $f(s,t)$ from $f(s_1, t)$ and $b(s_1, s)$ is the regression coefficient to predict $g(s,t+\tau)$ from $f(s_1, t)$. The spatial patterns in (8.6) are thus: $e_1(s)= a(s_1, s)$ and $d_1(s)= b(s_1, s)$. Note that $b(s_1, s)$ is proportional to the correlation defined in Eq(8.5) and used in Eq (8.7). We then seek the 2nd point in the once reduced data sets $f^{reduced}(s, t) = f(s, t) - a(s_1, s) f(s_1, t)$, and $g^{reduced}(s, t+\tau) = g(s, t+\tau) - b(s_1, s) f(s_1, t)$, to find s_2 etc. This procedure has many of the properties of CCA, specifically the identities (CCA-1), (CCA-2) and (CCA-3/3a), the latter with cor(m)=1 for all modes. (Oddly, EOT2 actually 'beats' CCA on producing the highest correlation between the time series.) EOT2 has at least two notions of relevance, the EV in data set f, and the EV in data set g. The latter is what is maximized, albeit under the constraint that we use a single time series of f at one point in space (rather than linear combinations of f at various points). There does not appear to be a particular need for the explained SC, after all the target of the prediction is EV in g.

Making a forecast of g is easy. For the first mode we need the observation of f at s_1, then multiply by $b(s_1, s)$. Subsequent modes are similar, but f has to be m-1 times reduced for the mth mode.

The reader will not be surprised that there is an 'alternative' lagged covariance matrix given by

$$c^a_{ij} = \Sigma\, f(s, t_i)\, g(s,t_j+\tau) / n_s \qquad (8.5b)$$

where summation is in space. Here we consider inner products of maps of fields f and g at times t_i and $t_j+\tau$. At first sight this definition is possible only if the domain and gridpoints for f and g are the same. However, this discrepancy is resolved by first executing EOFs on f and g individually and thinking of s in (8.5b) as the mode number. We now pick the one f map at time t_i which maximizes the variance explained in g, an expression similar to (8.7) but reversing the roles of time and space. This single map then acts as $e_1(s)$ for f and $d_1(s)$ for g. There are two time series, which are regression coefficients $a(t_1, t_i)$ to predict $f(s, t_i)$ from $f(s, t_1)$ and $b(t_1, t_i)$ to predict $g(s, t_i+\tau)$ from $f(s, t_1)$. This alternative EOT2 route leads to the expansion (8.6) and (8.6a) with the properties (SVD-1) and (SVD-2) but not (SVD-3). The alternative EOT2 has again two notions of relevance, the EV in data set f, and the EV in data set g. The latter is not only what is maximized[11], but is the purpose of the regression.

The two EOT versions that closely bracket CCA (regular EOT2) and SVD (alternative EOT2) come with either 2 maps and one time series (nearest CCA) or one map and two time series (nearest SVD). From this it appears that SVD is subject to more orthogonality constraints than CCA - after all (CCA-3) follows trivially when there is only one time series to begin with, but (SVD-3) does not follow automatically from having a single map (d=e).

Note that when admitting too many modes CCA/SVD goes in the direction of multiple linear regression. Obviously, truncation is necessary for reaping the benefits of regression at the pattern level.

Much information about SVD and CCA can be found in Bretherton et al(1992), Newman and Sareshmukh(1997) and Zwiers and von Storch(1999). Wilks(1995) provides a good discussion of CCA.

CCA was not used much in meteorology until Barnett and Preisendorfer(1987). The main methodological twist in their paper is a prefiltering step where both f and g are truncated to just a few EOFs before calculating C. (Moreover, the EOF associated time series are standardized, as in a version of the Mahalanobis norm (Stephenson 1997)) The prefiltering greatly reduces CCA's susceptibility to noise. The prefiltering also makes the practical difference between SVD and CCA in many instances very small. Additionally Barnett and Preisendorfer(1987) applied their adjusted

---

[11] under the constraint that we use maps of f at one point in time (rather than linear combinations at various times).

CCA to the seasonal forecast and had the predictor data set cover four antecedent seasons. This method and this particular predictor lay-out has been popularized by Barnston(1994) and his work found short-term climate prediction application on nearly all continents (Johansson et al (1998) for Europe; Thiaw et al(1999) for Africa; Hwong et al(2001) for Korea, Shabbar and Barnston(1996) for Canada, He and Barnston(1996) for tropical Pacific Islands and Barnston and Smith(1996) for the whole globe). While SVD is often mentioned in one breath with CCA, and widely used in research (Waliser et al 1999; Wu and Dickinson 2005) there appear to be far fewer real-time forecast applications based on SVD. CCA is also applied as a method to correct errors in GCM predictions (Smith and Livezey 1999; Tippett et al, 2005)

As a diagnostic tool SVD or CCA may be as difficult to use as EOF, i.e. the patterns in the predictor and predictand data set may or may not be revealing the underlying physics. Plenty of examples of patterns are found in Barnston(1994). Newman and Sardeshmukh (1997) show the failure (to a certain extent) of SVD to discover that vorticity and streamfunction are linear transforms of each other. Zwiers and Von Storch(1999) also provide several examples.

We spent some paragraphs explaining SVD, CCA etc because so much of the modern empirical work is along these lines. Regression on the pattern level is thought to take away the arbitrariness of correlating everything with everything. Although methodological details are hotly debated sometimes, the other choices may be more important than the exact method. For instance, which predictors, how far back in time, how many time levels, the domain for predictors and predictands, pre-filtering, truncation etc, may be more important than the exact CCA vs SVD method. The CCA at CPC and CDC, identically the same method, often give conflicting tropical Pacific SST forecasts. While we presented the above material as a strictly separated predictor f and predictand g, keep in mind that the data sets may be combined, i.e. fields of the predictand at an earlier time may be appended to f in order to forecast g. CCA has been used at both CPC and CDC for real time seasonal prediction; skill levels are at best (short lead JFM seasonal T&P) 0.3 - 0.35 correlation nationwide with regional variations that reflect the large impact of ENSO

(Barnston 1994; Quan et al 2005). The CCA modes suggest lesser influences from other tropical areas and mid-latitude oceans as well.

8.7.3 LIM, POP and Markov

Somewhat similar to CCA and SVD are the linear inverse model (LIM) and principal oscillation patterns(POP). The similarity is in the central role of the lagged covariance matrix as in (8.5), evaluated from data. However, both POP and LIM try to generalize the results for lag τ to all other lags by assuming an underlying theory. Following the Winkler et al(2001) notation one may assume a linear model given by

$$d\,x/dt = L\,x + R \qquad (8.8)$$

where x is the retained scales state vector, L is a linear operator and R is random forcing due to unresolved scales (possibly with structure in space). Vector x would for instance be a combination of data sets f and g. The solution to (8.8) is

$$x(t+\tau) = \exp(L\,\tau)\,x(t) + R', \qquad (8.9)$$

where R' depends on the history of R. The operator L can be determined from data at a chosen lag τ_0, i.e. we evaluate C for lag τ_0. L is given by $C(\tau_0)\,C^{-1}(\tau{=}0)$, see Winkler et al (2001) for detail. The forecast for any lag τ is given by the first term in (8.9). The forecasts for τ_0 would, everything else being the same, be close to CCA's. But an analytical flavor is added because time evolution is implied. Moreover, it is possible to calculate the eigenvectors of the asymmetric L once and for all - they are structures evolving in time, and ultimately damped. By knowing the projection of the current initial state onto the known eigenvectors of L, the forecast can be made analytically and can be interrogated for diagnostic purposes, such as in deriving the optimal structure to produce an El Nino pattern 10 months later (Penland and Sardeshmukh 1995). This is similar to what we presented for CA (section 7.6), although CA has additional growth due to unstable normal modes.

Several examples of POP, including for MJO forecasts, are given in Zwiers and von

Storch(1999). Winkler et al's (2001) application is in the week2 forecast, while Penland pioneered LIM for seasonal SST forecasts, both in the Pacific (Penland and Magorian 1993) and Atlantic (Penland and Matrosova 1998). In all cases C is calculated from EOF truncated data, but the degree of truncation varies wildly.

A straightforward method has been presented in Xue et al(2000). In this paper the discretized version of (8.8) is used: $x(t+\tau) = C(\tau)\, C^{-1}(\tau=0)\, x(t)$, i.e. given an initial state x(t) and $C(\tau)\, C^{-1}(\tau=0)$ as determined from data, the forecast for lead τ can be made. No linear model is assumed, so the calculation has to be done for each τ seperately, and nothing connects the forecasts at two different τ, except to the extent the data suggest. No modes are calculated, neither eigenmodes of L (as in LIM/POP, see Eq 8.9), or M (CCA) nor singular vectors of C (as in SVD). This cuts down on interpretation. The problem is handled as multiple linear regression, however after extremely heavy truncation using extended EOF in the input data. Xue et al(2000) use sea-level height, wind stress and SST to forecast the same (sea-level height, wind stress and SST) in the tropical Pacific which appears to be a wise choice, since the methods has worked well in real time. They call their method a 'Markov' (MRK) method. CCA, SVD, and LIM, POP and MRK have options in truncation both in preparing the input data, and in truncating the modes calculated from C, L or M.

8.8 Numerical Methods

This chapter would not be complete without discussing global atmosphere ocean models. Such models are the extension of NWP technology down into the ocean and consequently into processes with much longer time scales. It is generally assumed that most of the predictability of the seasonal climate resides in factors external to the atmosphere, such as the ocean, followed at some distance by land surface memory effects. In recent years, fully coupled atmosphere-ocean-land models have been developed to the point where they could start to make contributions to the real time seasonal forecast (Demeter May 2005 Tellus 57A#3 issue; Saha et al(2006) for the US).

This promise is realized only when plenty of hindcasts are available that allow both simple bias correction and more fancy calibration of pattern errors (Smith and Livezey 1999) and probability forecast adjustments. Raw model output can be very biased and needs postprocessing. In practice the number of hindcasts that any institutes can afford to run are limited by a) computer power and b) the period over which reliable initial states for land/ocean/atmosphere are available. Ocean analyses prior to 1981 may not be good enough as ICs or for verification. A 25 year period, multi-membered ensemble from each month is about the maximum that can be done (Saha et al 2006) at the current time. The initial states for land are a concern as well, but here too there has been an enormous effort (JGR 2003 issue, Fan et al 2006).

It is puzzling why comprehensive models, costing enormous resources, are only at par (roughly) with far simpler and empirical tools, and not much better. We come back to this conundrum in the final chapter. Saha et al(2006) show the NCEP coupled model (named CFS) to be on par with CCA, CA, Markov etc in NINO34 forecasts, and close to being at par with simpler methods for US T&P seasonal prediction. Van Oldenborgh et al (2003) have addressed the same question for the ECMWF model, and came to a similar conclusion.

8.9 Consolidation

Consolidation of multiple forecasts is necessary for a number of reasons. One can think of consolidation as the process of making the best possible official forecast out of a number of different forecast tools. This is a laudable goal for any weather service. While one may entertain the thought of consolidating subjectively, one very good reason for *objective consolidation* is that the supply of forecasts has become so large that no human (forecaster or user) is able to absorb the information in the available time, weigh their relative credibility, do justice to each component, and formulate the official forecast. And the problem is only getting worse. This is probably true for both short and long range forecasts, but here we address the latter. The idea of formal consolidation is at least as old as Thompson (1977), but development has been very slow.

We consider it self evident that a) one needs hindcasts in order to make an optimal consolidation and b) the real time forecasts need to be consistent in all respects with the hindcasts. But note that this practice has hardly been developed for dynamical models. CDC has undertaken a reforecast project (Hamill et al 2004) for the week2 forecast – this is a prototype of an atmosphere only reforecast. The new coupled ocean atmosphere model at NCEP (Saha et al 2006), the CFS, has a hindcast amounting to 3500 years of integration. In the case of Hamill et al. (2004) the purpose was calibration of the model in its own right, not consolidation. The CFS also needs to be combined with other methods like CCA. Other literature or practices are often employing AMIP runs (Gates et al 1999)for the hindcast of dynamical models (Peng et al 2002), but this does not apply in the real time setting. For statistical methods the idea of hindcasts comes more natural, has been common since about 1990, and is fairly well developed

Consolidation is difficult for many small reasons, but for two large reasons in particular. The first is a stunning lack of data vis-à-vis the number of coefficients that need fitting. Think of consolidation as CON = a*A + b*B +c*C….., where the capital letters refer to forecast tools, and the lower case coefficients need to be determined from hindcasts. Seasonal forecast tools like the CFS (Saha et al 2006) have only a 25 year record of hindcasts. This may sound like a lot, but in terms of seasonal means there is very little independent material to tune coefficients. While some empirical methods like CCA (Barnston 1994) may have 50 (or potentially 100) years, the data set still falls short by orders of magnitude if there are many methods/models to be combined. This is especially true when A, B, C to Z and all others are highly correlated. The 2nd major difficulty is low frequency climate change. The period 1981-2005 was much warmer over the US than the 60's and 70's. This major source of forecast skill (when verified against an 'old' normal) is accounted for in real time operations at CPC by the Optimal Climate Normals (OCN) method (Huang et al 1996). Dealing with this aspect retroactively on very limited data may be next to impossible.

By the nature of the consolidation methodology used one may see a stark distinction between two groups of consolidation activities, namely those working with point forecasts (it will be 66.0 F tomorrow) and those working with probabilities (next winter a 50% chance of the upper tercile). The distinction is not as absolute as it may seem because in applying linear regression to N point forecasts, one automatically obtains the root-mean-square error (rmse), which, to first approximation, serves as a standard deviation around the point forecast, i.e. linear regression results in a consolidated pdf as well. Whether the spread of forecasts, which varies on a case-by-case basis, can be used to improve upon the regression implied constant spread (based on rmse) remains an open question, but we note that Hamill et al. (2004) have concluded in the negative.

Consolidation is not new at NCEP. Leaving a long history of subjective consolidation aside, CPC has had a primitive consolidation of a few of its forecast tools. Van den Dool and Rukhovets(1994) designed a consolidation for the various members (at the time unequal members in terms of resolution, age) of the global model to support the 6-10 day forecast. Ever since Unger et al(1996), CPC has presented the official NWS SST forecast (Nino3.4 only) as a consolidation of in-house tools (CCA, CA, MRK, coupled model) – the exact method has evolved over time. In a sense 'ensemble' CCA (Mo, 2003) is also a consolidation. At the IRI a Bayesian method and a canonical variate method (Barnston et al 2003) are used to consolidate a large number of model forecasts from various centers around the world. The DEMETER project in Europe has been used to make a strong case in favor of multi-model ensembles (Hagedorn et al 2005; Stephenson et al 2005). Other research includes Kharin and Zwiers(2002) and Roulston and Smith(2003). CPC is working on consolidation as per ridge regression, see appendix 2.

Consolidation should be better than the single best participating tool, but for all the work to be done to determine optimal weights etc, the following should be kept in mind.

(1) The Consolidation will not be much better than the best individual tool if there is little independent information provided by the other tools

(2) The Consolidation might fail (in real time) if any of the hindcast data sets gives a

flawed impression of skill or if we cannot execute real time forecasts 100% consistent with the hindcasts. We do not have full control over these factors.

8.10 Other methods

In the above we have listed and described more or less all methods used operationally recently at institutions in the US (principally CPC, IRI and CDC) concerned with short-term climate prediction. While it is totally impossible to be complete, we here mention briefly some other methods and applications. It is not our intention to make an inventory of a toolbox, just tools that have been developed into methods for short-term climate prediction.

Multiple linear regression (MLR) has been used for ages. In Germany, under the leadership of Franz Baur, a truly ambitious MLR program was pursued in the early part of the twentieth century. The effort was handicapped naturally by the shortness of records at that time. MLR is dangerous in that chance correlations can be believed too easily by the hopeful, and that appears to have been widely the case. The over-reaction to the early mistakes was one of doubting anything empirical. Even the correlations associated with the Southern Oscillation (Berlage 1957), now considered essentially correct and seminal to understand ENSO, were considered suspect and lost their appeal when attention and resources were absorbed into the emerging NWP effort. Barnett (1981) and others re-introduced MLR and his work (and work by others, for instance Hastenrath and Greischar, 1993) includes a serious attempt to test regression on independent data, under cross validation etc, so as to largely avoid the errors made in the Baur era. Variations on the EOF theme may lead to any number of MLR approaches (Vautard et al 1999), and could include attempt to try non-linear regression (Kravtsov et al 2005).

There is a method called CLIPER, an acronym that suggests a combination of climatology and persistence, topics covered in 8.1 and 8.2. The history of CLIPER in hurricane prediction is to serve as a benchmark forecast to be improved upon by more complicated approaches. However, the CLIPER used in Landsea and Knaff (2000) and Knaff and Landsea(1997) for

Nino34 prediction also has non-local predictors, and predictors at more than one previous time, see sections 8.4 and 8.5. This is quite ambitious for a control method. CLIPER may well be among the better methods of forecasting Nino34.

Forecasts of PNA and NAO. The enormous attention for teleconnections and modes has given apparent reason to make PNA or NAO as such the predictand. A good example is Rodwell and Folland(2002) who concluded that a regression from May Atlantic SST to the NAO index in the next winter had more skill than SST predictors at other times. The quoted skill (in correlation) is about 0.4-0.5. A big problem with Atlantic predictors and the NAO are strong (inter)decadal variations. If relationships are strong at very low frequencies, there are very few degrees of freedom and claims of skill may have to be postponed until plentiful independent realizations are in. The UK Met Office has kept track of the performance of this scheme, see http://www.metoffice.com/research/seasonal/regional/nao/

A comprehensive verification of NAO and PNA forecasts at time scales ranging from daily to long lead seasonal and by many different methods is given by Johansson(2006).

There has recently been a shift in terms of predictands. Given the difficulty of forecasting precipitation attention has shifted to space-time integrated variables such as runoff, river flow, snowpack and soil moisture. These quantities have longer time scales, are easier to forecast (unless one invokes persistence as the control) than precipitation and are closer to application. Early examples include Cayan et al(1995) who essentially made ENSO composites for snowpack in the western US. Webster and Hoyos(2004) present a 10-30 day forecast scheme for major rivers in the South Asian subcontinent during the summer monsoon, based on a combination of wavelet analysis (Torrence and Compo 1998) and linear regression. Their predictors are chosen based on physical insights about the working of the Indian monsoon. Constructed analogue (chapter 7) has been applied to soil moisture over the US (Huang et al 1996) as predictor, to predict temperature, precipitation, evaporation and soil moisture over the US out to several months. While some skill was reported (Van den Dool et al 2003) it should be noted that the high skill for soil moisture (0.6 correlation) does not beat persistence. Van den Dool et al(2003) also contains a rock in the pond experiment where one places a round soil moisture anomaly somewhere in the US, anomalies being zero everywhere else, then let CA determine the motion of the soil moisture anomaly in the next several months in response to rainfall and evaporation anomalies. Variations on spectral methods, particularly singular spectrum analysis and maximum entropy methods, can be traced via a review article by Ghil et al(2002).

Checking the issues of the Experimental Long Lead Forecast Bulletin from its inception in 1992 to the present will provide access to a large array of methods experimented with by various authors. See http://www.iges.org/ellfb/home.html

There are good methods that have not been used in practice, although they could have been - this may be the accidental course of history up to this point. To delineate what is usable it may be best to list the requirement for any method to be usable The main requirements are about the method itself, the reproducibility of the calculations, and the unambiguous format of forecasts that need to be verified. The methods should be documented for the purpose of peer review - it also allows a potential user to judge whether this method could be any good. It should be possible, at least in principle, to reproduce a calculation (objective tools) and also to interrogate a method: why does Method X forecast such and such anomaly? After the fact it should be possible to make an unambiguous verification. An additional verification requirement is nowadays that of certified skill over a long enough testing period, under carefully designed conditions (to avoid erroneous skill estimates). These requirements rule out subjective prediction, or mostly subjective prediction. There has to be a method in the madness. Keeping these requirements in mind we now discuss methods that we do not recommend.

8.11 Methods not used

Some 15 years ago the author, in his capacity of Chief of Prediction, received a phone call from a New York Times reporter who first explained that a groundhog named Punxsutawney Phil had seen his shadow, so six more weeks of winter had to be expected. Then he asked: how does this compare to the official forecast. Suppressing a tendency to burst out in laughter, we politely answered the question. The reader of this book will probably agree immediately that sunshine in the early morning in one spot in Pennsylvania on February 2nd is an unlikely predictor of winter weather for the next 6 weeks. But how do we know so immediately? Based on the aforementioned requirements for usable methods the groundhog prediction fails on ambiguity (it is hard to verify the forecast as phrased), and also on method (this is unlikely to have skill). So we don't use Punxsutawney Phil in a professional environment. While it is impossible to prove categorically that any particular method is useless for any imaginable application we believe the same can be said about mapping the weather on the 12 days of Christmas onto the next 12 months, the hairiness of caterpillars in the fall, the size of beechnut crop etc. This is diversion and entertainment. Somewhat harder to dismiss off hand are the (ensemble of) farmer's almanacs. Persistent questions forced university researchers Walsh and Allen(1984) to attempt verification. Occasionally, during a long warm spell, one can hear a weatherman on TV say: 'things have to average out' so as to indicate that a compensating cold spell is unavoidable. This compensation idea is actually wrong. If it was true anti-persistence at a certain lag would prevail. A variant can be heard by economists commenting on the stock market: "what goes up must come down". These are truisms without any forecast skill (in meteorology) over a reasonable control. (Economic forecasts may come true when people believe the economist's truism and sell off their stock.)

A special topic is that of solar influences. While it makes perfect physical sense to assume that variations in solar output directly affect the atmosphere's temperature, we now know, from post 1979 satellite observations, that the solar constant varies only by about 0.1% over the course of the 11 year sunspot cycle. This may be detectable in global mean temperature, but is probably too small to be detected as a local effect. Possible amplification may exist via larger changes in ultraviolet radiation, and the stratospheric chemistry response. For more than a century before 1979 researchers have given the solar-weather connection a bad name by correlation everything with the 11 year cycle (known from sunspots). Even now anything solar as a predictor gets easily dismissed (over-reaction to past mistakes). There is a small probability this could change seriously (Lean and Rind 1999).

This section's title "Methods not used" does not quite cover the last example in this paragraph. In this case the method itself is fine, but its use and interpretation is debatable. In the distant past, the search of cycles (any cycles) was a prominent research activity. Even without identifying specific cycles as credible some methods involved Fourier analysis of very long time series, and the forecast for Δt ahead would be the extrapolation of all harmonics over Δt. This makes sense for periodic components (the atmospheric and oceanic tides), which we deleted from consideration from the outset as too easy to forecast (see start of this chapter). Fourier analysis followed by extrapolation to forecast anomalies in a chaotic atmosphere makes little sense (unless we have overlooked periodic components). Still this method is tried off and on, occasionally with modern twists such as EOFs etc.

Appendix 1 Some practical space-time continuity requirements

Many researchers may feel that maximizing the skill of a method is what is needed most. This can be done by wise choices about the prediction method, the predictor/predictand data sets and a cautious approach about cross validation. However, in practice there may be requirements that make the skill sub-optimal - these requirements are hard to deal with in research. We give a few examples.

1. Variation of K in OCN. In 8.3 we described how OCN is done. In principle this yields an optimal K value for each location, and every rolling season. However, because the U(K) may be a very flat function, and or have several minima the optimal K at nearby stations could be quite

different. For instance when K=2 in Washington DC and K=18 in Baltimore (60 km apart), very different OCN forecasts in real time could be the result. In order to avoid such irregularities on a national map, the function U(K) is actually minimized by summing over all locations and all seasons (as well as years). K=10 for temperature was derived that way. For a party that has local interests only, and is not concerned about inconsistencies in space the optimal K value (varying with season) may be better.

2. MLR inconsistencies. In a similar vain, multiple linear regression used at CPC (Unger personal communication) leads to inconsistencies because the best prediction equations for two nearby predictands are derived without regard for each other, can be quite different in the choice of predictors and may lead to spatial variation in a real time forecast that is impossible. It takes constraint on the part of the real time forecaster to decide what to believe. Regression at the pattern level largely avoids these problems, and may be preferred for that reason alone.

3. Consolidation of methods. When the weights assigned to methods are irregular, it could happen that certain models are 'coming in and going out' of the consolidation as a function of lead. This strikes the real time forecaster as an unlikely scenario. Measures should be taken to avoid that (by merging leads, points in space and neighboring initial conditions). Negative weights for forecasts also seem absurd. We assume that every reputable center does the best possible job and nobody has significant negative skill. (If they knew how to consistently give the anomaly opposite to observed they might as well give us the correct forecast.) Theoretically negative weights are possible but it is hard to sell a forecast for a positive anomaly in location X on the ground that world famous model A is forecasting a negative anomaly (and is always wrong; or gets a negative weight in a field of co-linear forecasts).

Appendix 2: Consolidation by Ridge Regression

For US T&P we consider to implement a version of Ridge Regression as the consolidation method. It is assumed that this will function even with an 'overload' of participating methods and short data sets – and this is the key consideration. The solution will be kept sane by pulling it slightly in the direction of a simpleminded approach based on just the skill of each method. Since skill ought to be positive, negative weights should never be assigned to any method. Space dependence of weights may be possible to some degree.

Definitions.

Let **A**, **B** and **C** be three[12] forecast methods with a hindcast history 1981-2003. **A** is shorthand for **A** (year, initial month, lead, space) or **A** (year, target month, lead, space). Stratification by month is customary, so **A** (y, l, s) suffices in the notation below, where y is 1981 to 2003, lead=1, 6(13), space (s) could be gridpoints NH (for example) or 102 super Climate Divisions in US. The matching observations are **O** (y, l, s). The inner product is defined by:

$$\mathbf{AB} = \Sigma\, \mathbf{A}(y, l, s) * \mathbf{B}(y, l, s) \qquad (8a.1)$$

where summation is over time y and (some or all of) space s. For simplicity we work with just three methods, but the derivation can be given for N methods.

In general we look for: CON(solidation) = a***A** + b***B** + c***C** (8a.2)

In a simple-minded solution a, b and c are proportional to the skill of methods **A**, **B** and **C**, i.e. proportional to **AO**, **BO** and **CO** (covariances), multiplied by 1/**AA** etc. In that case Σ a+b+c probably needs an additional constraint like a+b+c=1. a, b and c could be function of s, lead, initial (target) month. a, b and c should always be positive because we do not admit methods with negative skill (over the hindcast data set).

Full solution

While a simple-minded solution may be practical we actually seek the full optimal solution, taking into account both skill by methods and 'co-linearity' among methods:

$$\begin{array}{ccccc} \text{Matrix} & * & \text{vector} & = & \text{vector} \\ \begin{vmatrix} \mathbf{AA} & \mathbf{AB} & \mathbf{AC} \\ \mathbf{BA} & \mathbf{BB} & \mathbf{BC} \\ \mathbf{CA} & \mathbf{CB} & \mathbf{CC} \end{vmatrix} & * & \begin{vmatrix} a \\ b \\ c \end{vmatrix} & = & \begin{vmatrix} \mathbf{AO} \\ \mathbf{BO} \\ \mathbf{CO} \end{vmatrix} \end{array} \qquad (8a.3)$$

If co-linearity were zero, note a = **AO**/**AA** etc, the simple-minded solution. Also note there is no constraint on Σ a+b+c . The full solution takes co-linearity of methods into account. I.e. if A and B always give the same information they have to share the weight – they do not both get a high weight. The measure of co-linearity is given by the strength of the off-diagonal elements, relative to the main diagonal (AA, BB and CC).

[12] Three is just an example. Any number will do for explaining the process.

Including co-linearity is essential for the full solution, but problems arise when the co-linearity is too large, or when there is not enough data to estimate the co-linearity accurately. In either case the solution (a,b,c) to Eq (8a.3) may be unstable. In consolidation of seasonal forecasts there is FAR too little data to determine a,b,c…z, given the number of participating methods (quickly increasing).
Ridge regression is an amendment to the full solution to address this problem. One can stabilize the solution by adding small positive constants to the main diagonal of the matrix. This changes (8a.3) to (8a.4).

$$
\begin{array}{ccccc}
\text{Matrix} & * & \text{vector} & = & \text{vector} \\
\begin{vmatrix} \mathbf{AA}+\epsilon^2 & \mathbf{AB} & \mathbf{AC} \\ \mathbf{BA} & \mathbf{BB}+\epsilon^2 & \mathbf{BC} \\ \mathbf{CA} & \mathbf{CB} & \mathbf{CC}+\epsilon^2 \end{vmatrix} & * & \begin{vmatrix} a \\ b \\ c \end{vmatrix} & = & \begin{vmatrix} \mathbf{AO} \\ \mathbf{BO} \\ \mathbf{CO} \end{vmatrix}
\end{array} \qquad (8a.4)
$$

Even very small ϵ^2 can stabilize the worst possible matrices. Adding ϵ^2 to the main diagonal plays down the role of co-linearity ever so slightly, and drives the solution very slightly in the direction (but not exactly so) of a simple-minded solution. A 2nd layer of amplitude adjustment may be needed.

About Ridging
The oldest reference on ridging in the English literature is Tikhonov(1977) in translation, but this method may have been known since 1950 in Russia. The basic idea is to find a reasonable solution where there are more unknowns than equations. Nominally we have 3 equations and three unknowns in (8a.3), but when co-linearity is too large there may be effectively fewer than three equations. While (8a.3) minimize the rms difference between O and CON on a given data set, the ridge regression minimizes simultaneously Σ a*a+b*b+c*c. Ridging does this very effectively. The situation we encounter in consolidation is similar to data assimilation (Gandin 1965), where redundance (co-linearity) among observations to be assimilated is large. In the data assimilation context ϵ^2 relates to the (assumed) error in the observations. One could even embrace the situation as follows.

-) truncate forecast(obs) in EOF space
-) Now determine AA from filtered data….
-) Add ϵ^2 which is related to variance of unresolved EOFs
-) Controlled use of noise: off-diagonal elements unchanged.
-) solve the system.

In this context ϵ^2 has a real meaning, namely the variance of the unresolved EOFs.